На правах рукописи

Исаченко Роман Владимирович

Снижение размерности пространства в задачах декодирования сигналов

05.13.17 — Теоретические основы информатики

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва — 2021

Работа выполнена на Кафедре интеллектуальных систем Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)».

Научный руководитель: **Стрижов Вадим Викторович**
доктор физико-математических наук, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, отдел интеллектуальных систем, ведущий научный сотрудник.

Официальные оппоненты: **Чуличков Алексей Иванович**
доктор физико-математических наук, профессор, Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова», профессор кафедры математического моделирования и информатики физического факультета.

Зайцев Алексей Алексеевич
кандидат физико-математических наук, Автономная некоммерческая образовательная организация высшего образования «Сколковский институт науки и технологий», руководитель лаборатории в Центре по научным и инженерным вычислительным технологиям для задач с большими массивами данных.

Ведущая организация: Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики».

Защита состоится 6 февраля 2020 года в 13:00 на заседании диссертационного совета Д 002.073.05 при Федеральном исследовательском центре «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН) по адресу: 119333, г. Москва, ул. Вавилова, д. 40.

С диссертацией можно ознакомиться в библиотеке Федерального государственного учреждения Федеральный исследовательский центр «Информатика и управление» Российской академии наук и на сайте http://www.frccsc.ru/

Автореферат разослан декабря 2021 года.

И. о. ученого секретаря
диссертационного совета Д 002.073.05
д.т.н. И. А. Матвеев

Общая характеристика работы

Актуальность темы. В работе исследуется проблема снижения размерности пространства при решении задачи декодирования сигналов. Процесс декодирования заключается в восстановлении зависимости между двумя гетерогенными наборами данных. Прогностическая модель предсказывает отклик на входной исходный сигнал.

Исходное описание данных является избыточным. При высокой мультикорреляции в признаковом пространстве финальная прогностическая модель оказывается неустойчивой. Для построения простой, устойчивой модели применяются методы снижения размерности пространства (Motrenko: 2018, Chun: 2010, Mehmood: 2012) и выбора признаков (Katrutsa: 2017, Li: 2017).

В работе решается задача декодирования с векторной целевой переменной. Пространство целевых сигналов обладает избыточной размерностью. Методы снижения размерности, не учитывающие зависимости в целевом пространстве, не являются адекватными. При предсказании векторной целевой переменной анализируется структура целевого пространства. Предложены методы, которые учитывают зависимости как в пространстве исходных объектов, так и в пространстве целевой переменной. Предлагается отобразить пространства исходных и целевых сигналов в скрытые подпространства меньшей размерности. Для построения оптимальной модели предлагаются методы согласования скрытых пространств (Wold: 1975, Rosipal: 2005, Eliseyev: 2017). Предложенные методы позволяют учесть регрессионную компоненту между исходным и целевым сигналами, а также авторегрессионную компоненту целевого сигнала.

Методы снижения размерности пространства понижают размерность исходного пространства объектов, и, как следствие, сложность модели существенно снижается (Tipping: 1999, Hotelling: 1992). Алгоритмы снижения размерности находят оптимальные комбинации исходных признаков. Если число таких комбинаций существенно меньше, чем число исходных признаков, то полученное представление снижает размерность. Цель снижения размерности — получение наиболее репрезентативных и информативных комбинаций признаков для решения задачи.

Выбор признаков является частным случаем снижения размерности пространства (Katrutsa: 2015, 2017). Найденные комбинации признаков являются подмножеством исходных признаков. Таким образом отсеиваются шумовые неинформативные признаки. Рассматриваются два типа методов выбора признаков (Li: 2017, Rodriguez-Lujan: 2010, Friedman: 2001). Первый тип методов не зависит от последующей прогностической модели. Признаки отбираются на основе свойств исходных пространств, а не на основе свойств модели. Второй тип методов отбирает признаки с учётом знания о прогностической модели.

После нахождения оптимального представления данных с помощью снижения размерности, ставится задача нахождения оптимальной метрики в скрытом пространстве объектов (Wang: 2017, Davis: 2007, Kulis: 2012, Yang: 2006,

Weinberger: 2009). В случае евклидова пространства естественным выбором метрики оказывается квадратичная норма. Задача метрического обучения заключается в нахождении оптимальной метрики, связывающей объекты.

В качестве прикладной задачи анализируется задача построения нейрокомпьютерного интерфейса (Wolpaw: 2000, Allison: 2007). Цель состоит в извлечении информации из сигналов мозговой активности (Nagel: 2018, Zhang: 2020, Chiarelli: 2018). В качестве исходных сигналов выступают сигналы электроэнцефалограммы или электрокортикограммы. Целевым сигналом является траектория движения конечности индивидуума. Необходимо построить адекватную и эффективную модель декодирования исходного сигнала в целевой сигнал. Пространство частотных характеристик мозговых сигналов и авторегрессионное пространство целевых сигналов являются чрезвычайно избыточными (Eliseyev: 2011, 2013). Построение модели без учёта имеющихся зависимостей приводит к неустойчивости модели.

В диссертации решается задача декодирования с векторной целевой переменной. Для построения оптимальной модели декодирования сигналов предлагаются методы выбора согласованных моделей с проекцией в скрытое пространство. Исходные и целевые сигналы проецируются в пространство существенно меньшей размерности. Для связи проекций исходного и целевого сигналов предлагаются методы согласования. Рассматриваются гетерогенные наборы сигналов, природа источников измерений различны. Рассматриваются как линейные методы декодирования, так и их нелинейные обобщения. Доказаны теоремы об оптимальности предложенных методов выбора моделей.

Цели работы.

1. Исследовать свойства решения задачи декодирования сигналов с векторной целевой переменной.
2. Предложить методы снижения размерности пространства, учитывающие зависимости как в пространстве исходных сигналов, так и в целевом пространстве.
3. Предложить процедуру выбора признаков для задачи декодирования сигналов.
4. Исследовать свойства линейных и нелинейных моделей для решения поставленной модели. Получить теоретические оценки оптимальности моделей.
5. Провести вычислительные эксперименты для проверки адекватности предложенных методов.

Основные положения, выносимые на защиту.

1. Исследована проблема снижения размерности сигналов в коррелированных пространствах высокой размерности. Предложены методы декодирования сигналов, учитывающие зависимости как в исходном, так и в целевом пространстве сигналов.

2. Доказаны теоремы об оптимальности предлагаемых методов декодирования сигналов. Предлагаемые методы выбирают согласованные модели в случае избыточной размерности описания данных.
3. Предложены методы выбора признаков, учитывающие зависимости как в исходном, так и в целевом пространстве. Предложенные методы доставляют устойчивые и адекватные решения в пространствах высокой размерности.
4. Предложены нелинейные методы согласования скрытых пространств для данных со сложноорганизованной целевой переменной. Предложен метод выбора активных параметров для оптимизации нелинейной модели. Исследованы свойства предлагаемого метода.
5. Предложен алгоритм метрического обучения для временных рядов с процедурой их выравнивания.
6. Предложен ряд моделей для прогнозирования гетерогенных наборов сигналов для задачи построения нейрокомпьютерных интерфейсов. Проведены вычислительные эксперименты, подтверждающие адекватность моделей.

Методы исследования. Для достижения поставленных целей используются линейные и нелинейные методы регрессионного анализа. Для анализа временных рядов используются классические авторегрессионные методы. Для извлечения признаков используются частотные характеристики временного ряда. Для построения скрытого пространства используются линейные методы снижения размерности пространства, их нелинейные модификации, а также нейросетевые методы. Для выбора признаков наряду с классическими методами, используются методы, основанные на решении задачи квадратичного программирования. Для построения метрического пространства используются методы условной выпуклой оптимизации.

Научная новизна. Предложены методы построения моделей декодирования сигналов, учитывающие структуры пространств исходных и целевых переменных. Предложены методы проекции сигналов в скрытое пространство, а также процедуры согласования образов. Предложены методы выбора признаков с помощью квадратичного программирования. Предложен метод выбора активных параметров нелинейной модели с помощью выбора признаков. Предложены методы построения оптимального метрического пространства для задачи анализа временных рядов.

Теоретическая значимость. Доказаны теоремы об оптимальности предлагаемых моделей декодирования сигналов. Доказаны теоремы о корректности рассматриваемых согласованных моделей проекций в скрытое пространство. Доказаны теоремы о достижении точки равновесия для предлагаемых методов выбора признаков.

Практическая значимость. Предложенные в работе методы предназначены для декодирования множества временных рядов сигналов электрокортикограмм, а также нестационарных временных рядов; выбора оптимальных частотных характеристик сигналов; выбора наиболее информативных параметров модели; классификации и кластеризации временных рядов физической активности.

Степень достоверности и апробация работы. Достоверность результатов подтверждена математическими доказательствами, экспериментальной проверкой результатов предлагаемых методов на реальных данных, публикациями результатов в рецензируемых научных изданиях, в том числе рекомендованных ВАК. Результаты работы докладывались и обсуждались на следующих научных конференциях.

1. Р. В. Исаченко. Метрическое обучение в задачах мультиклассовой классификации временных рядов. *Международная научная конференция «Ломоносов»*, 2016.
2. R. G. Neychev, A. P. Motrenko, R. V. Isachenko, A. S. Inyakin, and V. V. Strijov. Multimodal forecasting multiscale time series in internet of things. *Международная научная конференция «11th International Conference on Intelligent Data Processing: Theory and Applications»*, 2016.
3. Р. В. Исаченко, И. Н. Жариков, и А. М. Бочкарёв. Локальные модели для классификации объектов сложной структуры. *Всероссийская научная конференция «Математические методы распознавания образов»*, 2017.
4. R. V. Isachenko and V. V. Strijov. Dimensionality reduction for multicorrelated signal decoding with projections to latent space. *Международная научная конференция «12th International Conference on Intelligent Data Processing: Theory and Applications»*, 2018.
5. Р. В. Исаченко, В. В. Стрижов. Снижение размерности в задаче декодирования временных рядов. *Международная научная конференция «13th International Conference on Intelligent Data Processing: Theory and Applications»*, 2020.

Публикации по теме диссертации. Основные результаты по теме диссертации изложены в 6 печатных изданиях, 5 из которых изданы в журналах, рекомендованных ВАК.

Структура и объем работы. Диссертация состоит из оглавления, введения, 6 глав, заключения, списка иллюстраций, списка таблиц, списка основных обозначений и списка литературы из 110 наименований. Основной текст занимает 121 страниц.

Личный вклад. Все приведенные результаты, кроме отдельно оговоренных случаев, получены диссертантом лично при научном руководстве д.ф.-м.н. В. В. Стрижова.

Основное содержание работы

Во **введении** обоснована актуальность диссертационной работы, сформулированы цели и методы исследования, обоснована научная новизна, теоретическая и практическая значимости полученных результатов.

В **главе 1** ставится общая задача декодирования временных рядов. Ставится задача построения оптимальной линейной регрессионной модели декодирования.

Пусть $\mathbb{X} \subset \mathbb{R}^n$ — пространство исходной переменной, $\mathbb{Y} \subset \mathbb{R}^r$ — пространство целевой переменной. Пусть задано множество объектов $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^m$, где $\mathbf{x}_i \in \mathbb{X}$ — исходный объект, $\mathbf{y}_i \in \mathbb{Y}$ — целевой объект.

Обозначим за $\mathbf{X} \in \mathbb{R}^{m \times n}$ матрицу исходной переменной, за $\mathbf{Y} \in \mathbb{R}^{n \times k}$ матрицу целевой переменной:

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]^\intercal = [\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_n]; \quad \mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_m]^\intercal = [\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_r].$$

Столбцы $\{\boldsymbol{\chi}_j\}_{j=1}^n$ матрицы $\mathbf{X}$ являются исходными признаками, столбцы $\{\boldsymbol{\nu}_j\}_{j=1}^r$ матрицы $\mathbf{Y}$ являются целевыми векторами.

Предполагается, что между исходным объектом $\mathbf{x}$ и целевым объектом $\mathbf{y}$ существует зависимость. Требуется построить прогностическую модель $\mathbf{f} : \mathbb{X} \to \mathbb{Y}$.

Задача восстановления регрессионной зависимости состоит в нахождении оптимальной модели $\mathbf{f}^*$ по заданным матрицам $\mathbf{X}$ и $\mathbf{Y}$:

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \mathcal{L}(\mathbf{f}, \mathbf{X}, \mathbf{Y}). \tag{1}$$

Для сужения пространства поиска моделей будем рассматривать параметрические модели $\mathbf{f}(\mathbf{x}, \boldsymbol{\Theta})$, где $\boldsymbol{\Theta}$ — *параметры модели*. Таким образом между объектами $\mathbf{x}$ и $\mathbf{y}$ существует зависимость вида

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\Theta}) + \boldsymbol{\varepsilon},$$

где $\mathbf{f}$ — параметрическая прогностическая модель, $\boldsymbol{\Theta}$ — параметры модели, $\boldsymbol{\varepsilon} \in \mathbb{R}^m$ — вектор регрессионных остатков.

Задача (1) сводится к задаче поиска оптимальных параметров

$$\boldsymbol{\Theta}^* = \arg\min_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\Theta}, \mathbf{X}, \mathbf{Y}).$$

Предположим, что зависимость $\mathbf{f}(\mathbf{x}, \boldsymbol{\Theta})$ линейная:

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\Theta}) + \boldsymbol{\varepsilon} = \boldsymbol{\Theta}^\intercal \mathbf{x} + \boldsymbol{\varepsilon}, \tag{2}$$

где $\boldsymbol{\Theta} \in \mathbb{R}^{n \times r}$ — матрица параметров модели.

Оптимальные параметры $\boldsymbol{\Theta}$ определяются минимизацией функции ошибки:

$$\mathcal{L}(\boldsymbol{\Theta}, \mathbf{X}, \mathbf{Y}) = \left\| \underset{m \times r}{\mathbf{Y}} - \underset{m \times n}{\mathbf{X}} \cdot \underset{r \times n}{\boldsymbol{\Theta}} \right\|_2^2 \to \min_{\boldsymbol{\Theta}}. \tag{3}$$

Решением (3) является следующая матрица:

$$\boldsymbol{\Theta} = (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{X}^\intercal \mathbf{Y}.$$

Если существует вектор $\boldsymbol{\alpha} \neq \mathbf{0}_n$ такой, что $\mathbf{X}\boldsymbol{\alpha} = \mathbf{0}_m$, то добавление $\boldsymbol{\alpha}$ к любому столбцу матрицы $\boldsymbol{\Theta}$ не меняет значение функции потерь $\mathcal{L}(\boldsymbol{\Theta}, \mathbf{X}, \mathbf{Y})$. В этом случае матрица $\mathbf{X}^\intercal \mathbf{X}$ близка к сингулярной и не обратима.

Задача декодирования сигналов состоит в восстановлении регрессионной зависимости (1) между наборами гетерогенных сигналов.

Пусть имеется два множества временных рядов $\mathcal{S}_\mathbf{x} = \{\mathbf{s}_\mathbf{x}^i\}_{i=1}^m$ и $\mathcal{S}_\mathbf{y} = \{\mathbf{s}_\mathbf{y}^i\}_{i=1}^r$, состоящие из m и r временных рядов соответственно. Первое множество $\mathcal{S}_\mathbf{x}$ является множеством временных рядов m исходных сигналов. Второе множество $\mathcal{S}_\mathbf{y}$ является множеством временных рядов r целевых сигналов. Каждый временной ряд $\mathbf{s} = (s_1, s_2, \ldots, s_T)$ является последовательностью измерений некоторый величины в течение времени.

Определение 1. *Временное представление* $\mathbf{x}_t = ([\mathbf{s}_\mathbf{x}^1]_t, \ldots, [\mathbf{s}_\mathbf{x}^m]_t) \in \mathbb{R}^m$ состоит из измерений временных рядов исходных сигналов в момент времени t. Аналогично временное представление $\mathbf{y}_t = ([\mathbf{s}_\mathbf{y}^1]_t, \ldots, [\mathbf{s}_\mathbf{y}^r]_t) \in \mathbb{R}^r$ состоит из измерений временных рядов целевых сигналов в момент времени t.

Определение 2. Определим *представление предыстории* длины h для момента времени t множества временных рядов исходных сигналов $\mathcal{S}_\mathbf{x}$ как совокупность представлений $\mathbf{X}_{t,h} = [\mathbf{x}_{t-h+1}, \ldots, \mathbf{x}_t]^\intercal \in \mathbb{R}^{h \times m}$. Аналогично определим представление предыстории длины h для момента времени t множества временных рядов целевых сигналов $\mathcal{S}_\mathbf{y}$ как совокупность представлений $\mathbf{Y}_{t,h} = [\mathbf{y}_{t-h+1}, \ldots, \mathbf{y}_t]^\intercal \in \mathbb{R}^{h \times r}$.

Определение 3. Определим *представление горизонта прогнозирования* длины p для момента времени t множества временных рядов исходных сигналов $\mathcal{S}_\mathbf{x}$ как совокупность представлений $\mathbf{X}_{t,p} = [\mathbf{x}_{t+1}, \ldots, \mathbf{x}_{t+p}]^\intercal \in \mathbb{R}^{p \times m}$. Аналогично определим представление горизонта прогнозирования длины p для момента времени t множества временных рядов целевых сигналов $\mathcal{S}_\mathbf{y}$ как совокупность представлений $\mathbf{Y}_{t,r} = [\mathbf{y}_{t+1}, \ldots, \mathbf{y}_{t+p}]^\intercal \in \mathbb{R}^{p \times r}$.

Определение 4. Прогностическая модель $\mathbf{f}_\mathbf{x}^{\mathrm{AR}} : \mathbb{R}^{h \times m} \to \mathbb{R}^{p \times m}$ является *авторегрессионной моделью*, которая по представлению предыстории $\mathbf{X}_{t,h}$ множества временных рядов исходных сигналов $\mathcal{S}_\mathbf{x}$ предсказывает представление горизонта прогнозирования $\mathbf{X}_{t,p}$ множества временных рядов исходных сигналов $\mathcal{S}_\mathbf{x}$. Аналогично вводится прогностическая модель $\mathbf{f}_\mathbf{y}^{\mathrm{AR}} : \mathbb{R}^{h \times r} \to \mathbb{R}^{p \times r}$ для множества целевых сигналов $\mathcal{S}_\mathbf{y}$.

Определение 5. Определим задачу *регрессионного декодирования* как задачу построения прогностической модели $\mathbf{f}_{\mathbf{xy}}^{\mathrm{R}} : \mathbb{R}^{h \times m} \to \mathbb{R}^{p \times r}$, которая по представлению предыстории $\mathbf{X}_{t,h}$ множества временных рядов исходных сигналов $\mathcal{S}_\mathbf{x}$ предсказывает представление горизонта прогнозирования $\mathbf{Y}_{t,p}$ множества временных рядов целевых сигналов $\mathcal{S}_\mathbf{y}$.

Определение 6. Общая *задача декодирования* состоит в построении прогностической модели $\mathbf{f_{xy}} : \mathbb{R}^{h_x \times m} \times \mathbb{R}^{h_y \times r} \to \mathbb{R}^{p \times r}$, которая по представлениям предыстории $\mathbf{X}_{t,h_x}$ и $\mathbf{Y}_{t,h_y}$ временных рядов исходных и целевых сигналов предсказывает представление горизонта прогнозирования $\mathbf{Y}_{t,r}$ временных рядов целевых сигналов.

Задача декодирования временных рядов декомпозируется на следующие подзадачи.

- Порождение признакового пространства. Процедура порождения признакового пространства может быть основана на экспертных знаниях или же являться моделью машинного обучения.
- Снижение размерности пространства или выбор признаков. Исходные временные ряды, а также порожденное признаковое пространство оказывается избыточным, что приводит к избыточности и неустойчивости модели.
- Построение модели. После нахождения оптимального низкоразмерного представления исходных данных ставится задача выбора оптимальной модели декодирования.

Метод главных компонент для задачи декодирования. Метод главных компонент (PCA) находит низкоразмерное представление матрицы $\mathbf{X} = \mathbf{TP}$, такое что новое представление $\mathbf{T} \in \mathbb{R}^{m \times l}$ содержит максимальную долю дисперсии исходной матрицы. При этом матрица отображения $\mathbf{P} \in \mathbb{R}^{l \times n}$ ($\mathbf{PP}^\intercal = \mathbf{I}$) содержит правые собственные вектора матрицы ковариаций $\mathbf{X}^\intercal\mathbf{X}$.

После нахождения матрицы отображения $\mathbf{P}$ задача (3) принимает вид

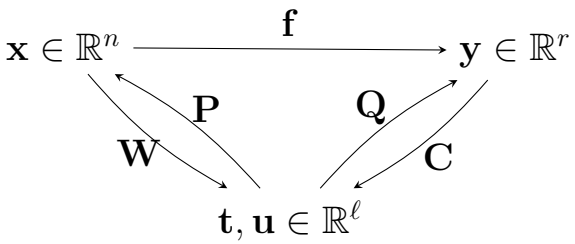
$$\mathcal{L}(\mathbf{B}, \mathbf{T}, \mathbf{Y}) = \left\| \underset{m \times r}{\mathbf{Y}} - \underset{m \times l}{\mathbf{T}} \cdot \underset{l \times r}{\mathbf{B}} \right\|_2^2 \to \min_{\mathbf{B}}.$$

Модель прогнозирования (2) в случае снижения размерности с помощью PCA принимает вид:

$$\mathbf{y} = \mathbf{B}^\intercal\mathbf{t} + \boldsymbol{\varepsilon} = \mathbf{B}^\intercal\mathbf{P}\mathbf{x} + \boldsymbol{\varepsilon} = \boldsymbol{\Theta}\mathbf{x} + \boldsymbol{\varepsilon}, \text{ где } \boldsymbol{\Theta} = \mathbf{B}^\intercal\mathbf{P}.$$

Метод частичных наименьших квадратов для задачи декодирования. Основным недостатком метода PCA является отсутствие учёта взаимосвязи между исходными признаками $\boldsymbol{\chi}_j$ и целевыми векторами $\boldsymbol{\nu}_j$. Метод частичных наименьших квадратов (PLS) проецирует матрицу исходных объектов $\mathbf{X}$ и матрицу целевых объектов $\mathbf{Y}$ в скрытое пространство малой размерностью l ($l < n$). Метод PLS находит в скрытом пространстве матрицы $\mathbf{T}, \mathbf{U} \in \mathbb{R}^{m \times l}$, которые лучше всего описывают исходные матрицы $\mathbf{X}$ и $\mathbf{Y}$. При этом PLS максимизирует ковариацию между столбцами матриц $\mathbf{T}$ и $\mathbf{U}$ соответственно. Принцип работы

метода показан на следующей коммутативной диаграмме:



Матрица исходных объектов $\mathbf{X}$ и матрица целевых объектов $\mathbf{Y}$ проецируются на скрытое пространство следующим образом:

$$\underset{m\times n}{\mathbf{X}} = \underset{m\times l}{\mathbf{T}} \cdot \underset{l\times n}{\mathbf{P}} + \underset{m\times n}{\mathbf{E}_\mathbf{x}} = \sum_{k=1}^{l} \underset{m\times 1}{\boldsymbol{\tau}_k} \cdot \underset{1\times n}{\mathbf{p}_k^\intercal} + \underset{m\times n}{\mathbf{E}_\mathbf{x}},$$

$$\underset{m\times r}{\mathbf{Y}} = \underset{m\times l}{\mathbf{U}} \cdot \underset{l\times r}{\mathbf{Q}} + \underset{m\times r}{\mathbf{E}_\mathbf{y}} = \sum_{k=1}^{l} \underset{m\times 1}{\boldsymbol{\nu}_k} \cdot \underset{1\times r}{\mathbf{q}_k^\intercal} + \underset{m\times r}{\mathbf{E}_\mathbf{y}}.$$

Здесь $\mathbf{T}$ и $\mathbf{U}$ — образы исходных матриц в скрытом пространстве, причём столбцы матрицы $\mathbf{T}$ ортогональны; $\mathbf{P}$ и $\mathbf{Q}$ — матрицы перехода; $\mathbf{E}_\mathbf{x}$ и $\mathbf{E}_\mathbf{y}$ — матрицы остатков. Метод PLS восстанавливает линейную зависимость между столбцами матриц $\mathbf{T}$ и $\mathbf{U}$

$$\mathbf{U} \approx \mathbf{TB}, \quad \mathbf{B} = \operatorname{diag}(\beta_k), \quad \beta_k = \boldsymbol{\nu}_k^\intercal \boldsymbol{\tau}_k / (\boldsymbol{\tau}_k^\intercal \boldsymbol{\tau}_k),$$

где $\{\boldsymbol{\tau}_k\}_{k=1}^l$, $\{\boldsymbol{\nu}_k\}_{k=1}^l$ — столбцы матриц $\mathbf{T}$ и $\mathbf{U}$ соответственно.

Метод решает следующую оптимизационную задачу:

$$\max_{\|\mathbf{p}\|_2=\|\mathbf{q}\|_2=1} [\operatorname{cov}(\mathbf{Xp}, \mathbf{Yq})^2] = \max_{\mathbf{p},\mathbf{q}} \frac{\mathbf{p}^\intercal \mathbf{X}^\intercal \mathbf{Y} \mathbf{q}}{\sqrt{\mathbf{p}^\intercal \mathbf{p}} \sqrt{\mathbf{q}^\intercal \mathbf{q}}}. \tag{4}$$

Канонический анализ корреляций для задачи декодирования. Оптимизационная задача канонического корреляционного анализа (CCA) похожа на оптимизационную задачу PLS (4) с той лишь разницей, что вместо максимизации ковариации CCA максимизирует корреляцию:

$$\max_{\|\mathbf{p}\|_2=\|\mathbf{q}\|_2=1} [\operatorname{corr}(\mathbf{Xp}, \mathbf{Yq})^2] = \max_{\mathbf{p},\mathbf{q}} \frac{\mathbf{p}^\intercal \mathbf{X}^\intercal \mathbf{Y} \mathbf{q}}{\sqrt{\mathbf{p}^\intercal \mathbf{X}^\intercal \mathbf{X} \mathbf{p}} \sqrt{\mathbf{q}^\intercal \mathbf{Y}^\intercal \mathbf{Y} \mathbf{q}}}.$$

Метод DeepCCA максимизирует корреляцию между представлениями, полученными на выходе нейросети:

$$\max_{\|\mathbf{p}\|_2=\|\mathbf{q}\|_2=1} \left[\operatorname{corr}(\mathbf{G}_\mathbf{x}(\mathbf{X}, \mathbf{W}_\mathbf{x}) \cdot \mathbf{p}, \mathbf{G}_\mathbf{y}(\mathbf{Y}, \mathbf{W}_\mathbf{y}) \cdot \mathbf{q})^2\right] =$$

$$= \max_{\mathbf{p},\mathbf{q}} \frac{\mathbf{p}^\intercal \mathbf{G}_\mathbf{x}(\mathbf{X}, \mathbf{W}_\mathbf{x})^\intercal \mathbf{G}_\mathbf{y}(\mathbf{Y}, \mathbf{W}_\mathbf{y}) \mathbf{q}}{\sqrt{\mathbf{p}^\intercal \mathbf{G}_\mathbf{x}(\mathbf{X}, \mathbf{W}_\mathbf{x})^\intercal \mathbf{G}_\mathbf{x}(\mathbf{X}, \mathbf{W}_\mathbf{x}) \mathbf{p}} \sqrt{\mathbf{q}^\intercal \mathbf{G}_\mathbf{y}(\mathbf{Y}, \mathbf{W}_\mathbf{y})^\intercal \mathbf{G}_\mathbf{y}(\mathbf{Y}, \mathbf{W}_\mathbf{y}) \mathbf{q}}}.$$

Здесь $\mathbf{G}_{\mathbf{x}}(\mathbf{X}, \mathbf{W}_{\mathbf{x}})$ и $\mathbf{G}_{\mathbf{y}}(\mathbf{Y}, \mathbf{W}_{\mathbf{y}})$ — нелинейные проекции исходных и целевых объектов. С использованием нейросетевых функций модель декодирования способна учитывать существенно нелинейные зависимости как в исходном пространстве, так и в целевом пространстве.

В **главе 2** приводится формальная постановка задачи построения согласованных моделей декодирования. Вводятся понятия скрытого пространства и процедуры согласования образов. Доказываются теоремы о выборе оптимальной модели декодирования.

Предположение 1. Пусть пространства $\mathbb{X}$ и $\mathbb{Y}$ имеют избыточную размерность. В простейшем случае такими многообразия могут являться вложениями или линейными подпространствами.

Определение 7. Назовём пространство $\mathbb{T} \subset \mathbb{R}^l$ *скрытым пространством* для пространства $\mathbb{X} \subset \mathbb{R}^n$ $(l \leqslant n)$, если существуют *функция кодирования* $\boldsymbol{\varphi}_{\mathbf{x}} : \mathbb{X} \to \mathbb{T}$ и *функция декодирования* $\boldsymbol{\psi}_{\mathbf{x}} : \mathbb{T} \to \mathbb{X}$, такие что

$$\text{для любого } \mathbf{x} \in \mathbb{X} \quad \text{существует } \mathbf{t} \in \mathbb{T} : \boldsymbol{\psi}_{\mathbf{x}}(\boldsymbol{\varphi}_{\mathbf{x}}(\mathbf{x})) = \boldsymbol{\psi}_{\mathbf{x}}(\mathbf{t}) = \mathbf{x}.$$

Аналогично введём определение *скрытого пространства* $\mathbb{U} \subset \mathbb{R}^s$ для целевого пространства $\mathbb{Y}$, *функции кодирования* $\boldsymbol{\varphi}_{\mathbf{y}} : \mathbb{Y} \to \mathbb{U}$ и *декодирования* $\boldsymbol{\psi}_{\mathbf{y}} : \mathbb{U} \to \mathbb{Y}$, такие что

$$\text{для любого } \mathbf{y} \in \mathbb{Y} \quad \text{существует } \mathbf{u} \in \mathbb{U} : \boldsymbol{\psi}_{\mathbf{y}}(\boldsymbol{\varphi}_{\mathbf{y}}(\mathbf{y})) = \boldsymbol{\psi}_{\mathbf{y}}(\mathbf{u}) = \mathbf{y}.$$

Образы матрицы исходных объектов $\mathbf{X}$ и матрицы целевых объектов $\mathbf{Y}$ в скрытых пространствах $\mathbb{T}$ и $\mathbb{U}$ имеют вид

$$\mathbf{T} = \boldsymbol{\varphi}_{\mathbf{x}}(\mathbf{X}) = [\mathbf{t}_1, \ldots, \mathbf{t}_n]^{\mathsf{T}} = [\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_l],$$
$$\mathbf{U} = \boldsymbol{\varphi}_{\mathbf{y}}(\mathbf{U}) = [\mathbf{u}_1, \ldots, \mathbf{u}_n]^{\mathsf{T}} = [\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_s].$$

Определение 8. Будем говорить, что скрытые пространства $\mathbb{T}$ и $\mathbb{U}$ являются *согласованными*, если существует *функция связи* $\mathbf{h} : \mathbb{T} \to \mathbb{U}$, такая что

$$\text{для любого } \mathbf{u} \in \mathbb{U} \quad \text{существует } \mathbf{t} \in \mathbb{T} : \mathbf{u} = \mathbf{h}(\mathbf{t}).$$

Предположение 2. Предположим, что в задаче прогнозирования (1) пространства $\mathbb{T}$ и $\mathbb{U}$ являются скрытыми для пространств $\mathbb{X}$ и $\mathbb{Y}$ соответственно. Предположим также, что для данных скрытых пространств $\mathbb{T}$ и $\mathbb{U}$ существует функция связи $\mathbf{h} : \mathbb{T} \to \mathbb{U}$. Тогда выполнено

$$\text{для любого } \mathbf{y} \in \mathbb{Y} \quad \text{существует } \mathbf{x} \in \mathbb{X} : \mathbf{y} = \boldsymbol{\psi}_{\mathbf{y}}(\mathbf{u}) = \boldsymbol{\psi}_{\mathbf{y}}(\mathbf{h}(\mathbf{t})) = \boldsymbol{\psi}_{\mathbf{y}}(\mathbf{h}(\boldsymbol{\varphi}_{\mathbf{x}}(\mathbf{x}))),$$

и общая схема задачи поиска согласованной модели декодирования принимает

вид следующей коммутативной диаграммы:

$$\begin{array}{ccc} \mathbb{X} \subset \mathbb{R}^n & \xrightarrow{\quad\mathbf{f}\quad} & \mathbb{Y} \subset \mathbb{R}^r \\ \boldsymbol{\varphi}_{\mathbf{x}} \big\uparrow\big\downarrow \boldsymbol{\psi}_{\mathbf{x}} & & \boldsymbol{\psi}_{\mathbf{y}} \big\uparrow\big\downarrow \boldsymbol{\varphi}_{\mathbf{y}} \\ \mathbb{T} \subset \mathbb{R}^\ell & \xrightarrow{\quad\mathbf{h}\quad} & \mathbb{U} \subset \mathbb{R}^s \end{array} \tag{5}$$

Определение 9. Согласно диаграмме (5), определим *согласованную* модель декодирования $\mathbf{f} : \mathbb{X} \to \mathbb{Y}$ как суперпозицию

$$\mathbf{f} = \boldsymbol{\psi}_{\mathbf{y}} \circ \mathbf{h} \circ \boldsymbol{\varphi}_{\mathbf{x}}.$$

Для поиска оптимальных параметров функций кодирования $\boldsymbol{\varphi}_{\mathbf{x}}$ и $\boldsymbol{\varphi}_{\mathbf{y}}$, декодирования $\boldsymbol{\psi}_{\mathbf{x}}$ и $\boldsymbol{\psi}_{\mathbf{y}}$, а также функции связи $\mathbf{h}$ ставится задача максимизации *функции согласования скрытых векторов*

$$g : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}, \quad g(\boldsymbol{\tau}, \boldsymbol{\nu}) \to \max_{\boldsymbol{\varphi}_{\mathbf{x}}, \boldsymbol{\varphi}_{\mathbf{y}}, \mathbf{h}}.$$

Каждая пара скрытых векторов $\boldsymbol{\tau}, \boldsymbol{\nu}$ ищется последовательно.

Метод главных компонент. Функции кодирования $\boldsymbol{\varphi}_{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}^l$ и декодирования $\boldsymbol{\psi}_{\mathbf{x}} : \mathbb{R}^l \to \mathbb{R}^n$ имеют вид

$$\boldsymbol{\varphi}_{\mathbf{x}}(\mathbf{X}) = \underset{m\times n}{\mathbf{X}} \cdot \underset{n\times l}{\mathbf{P}}^{\mathsf{T}}, \quad \boldsymbol{\psi}_{\mathbf{x}}(\mathbf{T}) = \underset{m\times l}{\mathbf{T}} \cdot \underset{l\times n}{\mathbf{P}},$$

где $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_l]^{\mathsf{T}}$, при этом $\mathbf{P}\mathbf{P}^{\mathsf{T}} = \mathbf{I}$.

Скрытые вектора $\boldsymbol{\tau}$ строятся так, чтобы выборочная дисперсия столбцов проекций матрицы $\mathbf{X}$ была максимальной:

$$\mathbf{p} = \underset{\|\mathbf{p}\|_2=1}{\arg\max}\, g(\boldsymbol{\tau}) = \underset{\|\mathbf{p}\|_2=1}{\arg\max}[\mathrm{var}(\boldsymbol{\tau})] = \underset{\|\mathbf{p}\|_2=1}{\arg\max}[\mathrm{var}(\mathbf{X}\mathbf{p})].$$

Метод PCA не согласует исходные переменные и целевые переменные. А именно метод PCA не находит функции кодирования $\boldsymbol{\varphi}_{\mathbf{y}}$ и декодирования $\boldsymbol{\psi}_{\mathbf{y}}$, а также функцию связи $\mathbf{h}$. При этом функция согласования скрытых векторов $g(\boldsymbol{\tau})$ зависит только от одного аргумента.

Метод частичных наименьших квадратов и канонический анализ корреляций. В методах PLS и CCA функции кодирования и декодирования имеют вид

$$\begin{aligned} &\boldsymbol{\varphi}_{\mathbf{x}}(\mathbf{X}) = \mathbf{X}\mathbf{W}, \quad \boldsymbol{\varphi}_{\mathbf{y}}(\mathbf{Y}) = \mathbf{Y}\mathbf{C}, \\ &\boldsymbol{\psi}_{\mathbf{x}}(\mathbf{T}) = \mathbf{T}\mathbf{P}^{\mathsf{T}}, \quad \boldsymbol{\psi}_{\mathbf{y}}(\mathbf{U}) = \mathbf{U}\mathbf{Q}^{\mathsf{T}}. \end{aligned}$$

Функция связи $\mathbf{h}$ имеет вид линейной модели, связывающей образы проекций в скрытом пространстве $\mathbf{u} = \mathbf{h}(\mathbf{t}) = \mathbf{B}^\intercal\mathbf{t}$. В данном случае схема декодирования (5) принимает вид следующей коммутативной диаграммы.

$$\begin{array}{ccc} \mathbb{X} \subset \mathbb{R}^n & \xrightarrow{\quad\mathbf{f}\quad} & \mathbb{Y} \subset \mathbb{R}^r \\ \mathbf{W}\big\uparrow\big\downarrow\mathbf{P} & & \mathbf{Q}\big\uparrow\big\downarrow\mathbf{C} \\ \mathbb{T} \subset \mathbb{R}^\ell & \xrightarrow{\quad\mathbf{B}\quad} & \mathbb{U} \subset \mathbb{R}^s \end{array}$$

Для метода PLS функция согласования скрытых векторов имеет вид $g(\boldsymbol{\tau}, \boldsymbol{\nu}) = \operatorname{cov}(\boldsymbol{\tau}, \boldsymbol{\nu})$, а для метода CCA: $g(\boldsymbol{\tau}, \boldsymbol{\nu}) = \operatorname{corr}(\boldsymbol{\tau}, \boldsymbol{\nu})$.

Нелинейный канонический анализ корреляций. Функции кодирования и декодирования являются нелинейными нейросетями вида

$$\begin{aligned} \mathbf{T} &= \boldsymbol{\varphi}_\mathbf{x}(\mathbf{X}) = \mathbf{W}_\mathbf{x}^L\sigma(\dots\sigma(\mathbf{W}_\mathbf{x}^2\sigma(\mathbf{X}\mathbf{W}_\mathbf{x}^1))\dots), \\ \mathbf{U} &= \boldsymbol{\varphi}_\mathbf{y}(\mathbf{Y}) = \mathbf{W}_\mathbf{y}^L\sigma(\dots\sigma(\mathbf{W}_\mathbf{y}^2\sigma(\mathbf{Y}\mathbf{W}_\mathbf{y}^1))\dots), \\ \mathbf{X} &= \boldsymbol{\psi}_\mathbf{x}(\mathbf{T}) = \mathbf{W}_\mathbf{t}^L\sigma(\dots\sigma(\mathbf{W}_\mathbf{t}^2\sigma(\mathbf{T}\mathbf{W}_\mathbf{t}^1))\dots), \\ \mathbf{Y} &= \boldsymbol{\psi}_\mathbf{y}(\mathbf{U}) = \mathbf{W}_\mathbf{u}^L\sigma(\dots\sigma(\mathbf{W}_\mathbf{u}^2\sigma(\mathbf{U}\mathbf{W}_\mathbf{u}^1))\dots). \end{aligned}$$

Каждая нейросеть является суперпозицией последовательных умножений на матрицы параметров и применения поэлементных функций активаций.

Требуется найти такие параметры, при которых функция согласования g достигает своего максимума:

$$g(\boldsymbol{\tau}, \boldsymbol{\nu}) \to \max_{\mathbf{W}},$$

где $\mathbf{W} = \left\{\mathbf{W}_\mathbf{x}^i, \mathbf{W}_\mathbf{y}^i, \mathbf{W}_\mathbf{t}^i, \mathbf{W}_\mathbf{u}^i\right\}_{i=1}^L$.

При использовании в качестве функции согласования корреляции между проекциями $g(\boldsymbol{\tau}, \boldsymbol{\nu}) = \operatorname{corr}(\boldsymbol{\tau}, \boldsymbol{\nu})$ частная производная функции согласования по первому аргументу принимает вид

$$\frac{\partial\mathbf{g}(\mathbf{T}, \mathbf{U})}{\partial\mathbf{T}} = \frac{1}{m-1}\left(\boldsymbol{\Sigma}_1^{-1/2}\mathbf{U}\mathbf{V}^\intercal\boldsymbol{\Sigma}_2^{-1/2}\mathbf{U} - \boldsymbol{\Sigma}_1^{-1/2}\mathbf{U}\mathbf{D}\mathbf{V}^\intercal\boldsymbol{\Sigma}_1^{-1/2}\right),$$

где $\mathbf{U}, \mathbf{D}, \mathbf{V} = \operatorname{SVD}(\boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_2^{-1/2}$, $\boldsymbol{\Sigma}_1 = \frac{1}{m-1}\mathbf{T}\mathbf{T}^\intercal$, $\boldsymbol{\Sigma}_2 = \frac{1}{m-1}\mathbf{U}\mathbf{U}^\intercal$, $\boldsymbol{\Sigma}_{12} = \frac{1}{m-1}\mathbf{T}\mathbf{U}^\intercal$. Аналогичное выражение имеет частная производная по второму аргументу. Полученное выражение для градиента позволяет построить эффективный алгоритм для решения задачи с использованием градиентных методов оптимизации.

Алгоритм PLS итеративно на каждом из l шагов вычисляет по одному столбцу $\boldsymbol{\tau}_k$, $\boldsymbol{\nu}_k$ матриц $\mathbf{T}$, $\mathbf{U}$ и по одной строке $\mathbf{p}_k$, $\mathbf{q}_k$ матриц $\mathbf{P}$, $\mathbf{Q}$ соответственно.

Утверждение 1. Максимизация ковариации между векторами $\boldsymbol{\tau}_k$ и $\boldsymbol{\nu}_k$ сохраняет дисперсию столбцов матриц $\mathbf{X}$ и $\mathbf{Y}$ и учитывает их линейную зависимость.
Утверждение 2. Вычисленные вектора $\mathbf{w}_k$ и $\mathbf{c}_k$ с помощью итеративной процедуры обновления:

$$\boldsymbol{\tau}_k := \frac{\mathbf{X}_k\mathbf{w}_k}{\|\mathbf{w}_k\|}, \quad \mathbf{w}_k := \frac{\mathbf{X}_k^{\mathsf{T}}\boldsymbol{\nu}_{k-1}}{\boldsymbol{\nu}_{k-1}^{\mathsf{T}}\boldsymbol{\nu}_{k-1}}; \tag{6}$$

$$\boldsymbol{\nu}_k := \frac{\mathbf{Y}_k\mathbf{c}_k}{\|\mathbf{c}_k\|}, \quad \mathbf{c}_k := \frac{\mathbf{Y}_k^{\mathsf{T}}\boldsymbol{\tau}_k}{\boldsymbol{\tau}_k^{\mathsf{T}}\boldsymbol{\tau}_k}. \tag{7}$$

будут собственными векторами матриц $\mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k\mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k$ и $\mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k\mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k$, соответствующими максимальным собственным значениям.

$$\mathbf{w}_k \propto \mathbf{X}_k^{\mathsf{T}}\boldsymbol{\nu}_{k-1} \propto \mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k\mathbf{c}_{k-1} \propto \mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k\mathbf{Y}_k^{\mathsf{T}}\boldsymbol{\tau}_{k-1} \propto \mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k\mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k\mathbf{w}_{k-1},$$

$$\mathbf{c}_k \propto \mathbf{Y}_k^{\mathsf{T}}\boldsymbol{\tau}_k \propto \mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k\mathbf{w}_k \propto \mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k\mathbf{X}_k^{\mathsf{T}}\boldsymbol{\nu}_{k-1} \propto \mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k\mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k\mathbf{c}_{k-1},$$

где символ $\propto$ означает равенство с точностью до мультипликативной константы.

Утверждение 3. Вычисленные вектора $\boldsymbol{\tau}_k$ и $\boldsymbol{\nu}_k$ с помощью итеративной процедуры обновления (6), (7) обладают максимальной ковариацией $\operatorname{cov}(\boldsymbol{\tau}, \boldsymbol{\nu})$.

Для перехода на следующий шаг необходимо вычесть из матриц $\mathbf{X}_k$ и $\mathbf{Y}_k$ одноранговые аппроксимации $\boldsymbol{\tau}_k\mathbf{p}_k^{\mathsf{T}}$ и $\boldsymbol{\tau}_k\mathbf{q}_k^{\mathsf{T}}$

$$\mathbf{X}_{k+1} = \mathbf{X}_k - \boldsymbol{\tau}_k\mathbf{p}_k^{\mathsf{T}} = \mathbf{X} - \sum_k \boldsymbol{\tau}_k\mathbf{p}_k^{\mathsf{T}},$$

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k - \boldsymbol{\tau}_k\mathbf{q}_k^{\mathsf{T}} = \mathbf{Y} - \sum_k \boldsymbol{\tau}_k\mathbf{q}_k^{\mathsf{T}}.$$

Теорема 1. В случае линейных функций декодирования $\boldsymbol{\psi}_{\mathbf{x}}(\mathbf{T}) = \mathbf{TP}$, $\boldsymbol{\psi}_{\mathbf{y}}(\mathbf{U}) = \mathbf{UQ}$ и функции согласования $g(\boldsymbol{\tau}, \boldsymbol{\nu}) = \operatorname{cov}(\boldsymbol{\tau}, \boldsymbol{\nu})$ параметры

$$\boldsymbol{\Theta} = \mathbf{W}(\mathbf{P}^{\mathsf{T}}\mathbf{W})^{-1}\mathbf{BQ}$$

являются оптимальными для модели (2).

Финальная модель является линейной, низкоразмерной в скрытом пространстве. Это снижает избыточность данных и повышает стабильность модели.

Пусть $\mathbf{f}_1(\mathbf{x}_1, \boldsymbol{\Theta}_1)$, $\mathbf{f}_2(\mathbf{x}_2, \boldsymbol{\Theta}_2)$ — линейные модели декодирования сигналов.
Определение 10. Назовём *аддитивной суперпозицией* моделей декодирования модель (2) вида

$$\mathbf{Y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\Theta}) + \boldsymbol{\varepsilon} = \mathbf{f}_1(\mathbf{x}_1, \boldsymbol{\Theta}_1) + \mathbf{f}_2(\mathbf{x}_2, \boldsymbol{\Theta}_2) + \boldsymbol{\varepsilon} = \boldsymbol{\Theta}_1^{\mathsf{T}}\mathbf{x}_1 + \boldsymbol{\Theta}_2^{\mathsf{T}}\mathbf{x}_2 + \boldsymbol{\varepsilon}, \tag{8}$$

где объект $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2] \in \mathbb{R}^n$ состоит из двух подвекторов $\mathbf{x}_1 \in \mathbb{R}^k$, $\mathbf{x}_2 \in \mathbb{R}^{n-k}$.

Утверждение 4. Оптимальная матрица параметров $\mathbf{\Theta}$ для модели (8), доставляющая минимум функции ошибки (3), имеет вид:

$$\mathbf{\Theta}_1 = (\mathbf{X}_1^\intercal \mathbf{M}_{\mathbf{X}_2} \mathbf{X}_1)^{-1} \mathbf{X}_1^\intercal \mathbf{M}_{\mathbf{X}_2} \mathbf{Y},$$
$$\mathbf{\Theta}_2 = (\mathbf{X}_2^\intercal \mathbf{M}_{\mathbf{X}_1} \mathbf{X}_2)^{-1} \mathbf{X}_2^\intercal \mathbf{M}_{\mathbf{X}_1} \mathbf{Y},$$

где $\mathbf{M}_{\mathbf{X}_1} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}_1})$, $\mathbf{M}_{\mathbf{X}_2} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}_2})$, $\mathbf{P}_{\mathbf{X}_1} = \mathbf{X}_1(\mathbf{X}_1^\intercal \mathbf{X}_1)^{-1}\mathbf{X}_1^\intercal$, $\mathbf{P}_{\mathbf{X}_2} = \mathbf{X}_2(\mathbf{X}_2^\intercal \mathbf{X}_2)^{-1}\mathbf{X}_2^\intercal$.

Утверждение 5. Оптимальная подматрица $\mathbf{\Theta}_2$ в модели (8) является решением задачи регрессии

$$\|\mathbf{Y}_1 - \mathbf{X}_{21}\mathbf{\Theta}_2\|^2 \to \min_{\mathbf{\Theta}_2},$$

где $\mathbf{Y}_1 = \mathbf{M}_{\mathbf{X}_1}\mathbf{Y}$, $\mathbf{X}_{21} = \mathbf{M}_{\mathbf{X}_1}\mathbf{X}_2$.

Утверждение 6. Если в задаче (8) $\operatorname{span}(\mathbf{X}_1) \cap \operatorname{span}(\mathbf{X}_2) = \varnothing$, то есть столбцы матрицы $\mathbf{X}_1$ ортогональны столбцам матрицы $\mathbf{X}_2$, то $\mathbf{\Theta}_2$ является решением задачи регрессии

$$\|\mathbf{Y} - \mathbf{X}_2\mathbf{\Theta}_2\|^2 \to \min_{\mathbf{\Theta}_2}.$$

Утверждение 7. Ошибка аддитивной суперпозиции моделей не превышает ошибки каждой из отдельных моделей

$$\mathcal{L}(\mathbf{\Theta}^*, \mathbf{X}, \mathbf{Y}) \leqslant \mathcal{L}(\mathbf{\Theta}_1, \mathbf{X}_1, \mathbf{Y}),$$
$$\mathcal{L}(\mathbf{\Theta}^*, \mathbf{X}, \mathbf{Y}) \leqslant \mathcal{L}(\mathbf{\Theta}_2, \mathbf{X}_2, \mathbf{Y}).$$

Утверждение 8. Пусть для аддитивной суперпозиции моделей (8) выполнены следующие условия

$$\mathbf{Y} \neq \mathbf{P}_{\mathbf{X}_2}\mathbf{Y}, \quad \mathbf{X}_1 \neq \mathbf{P}_{\mathbf{X}_2}\mathbf{X}_1, \quad \mathbf{Y}^\intercal \mathbf{M}_{\mathbf{X}_2}\mathbf{X}_1 \neq 0.$$

Тогда выполнено строгое неравенство

$$\mathcal{L}(\mathbf{\Theta}^*, \mathbf{X}, \mathbf{Y}) < \mathcal{L}(\mathbf{\Theta}_2, \mathbf{X}_2, \mathbf{Y}).$$

Глава 3 посвящена задаче выбора признаков для задачи декодирования сигналов. Задача выбора признаков заключается в поиске оптимального подмножества $\mathcal{A} \subset \{1, \ldots, n\}$ индексов признаков среди всех возможных $2^n - 1$ вариантов. Существует взаимооднозначное отображение между подмножеством $\mathcal{A}$ и булевым вектором $\mathbf{a} \in \{0,1\}^n$, компоненты которого указывают, выбран ли признак. Для нахождения оптимального вектора $\mathbf{a}$ введем функцию ошибки выбора признаков $S(\mathbf{a}, \mathbf{X}, \mathbf{Y})$:

$$\mathbf{a} = \arg\min_{\mathbf{a}' \in \{0,1\}^n} S(\mathbf{a}', \mathbf{X}, \mathbf{Y}). \tag{9}$$

Целью выбора признаков является построение функции $S(\mathbf{a}, \mathbf{X}, \mathbf{Y})$.

Для решения задачи (9) применяется релаксация задачи (9) к непрерывной области определения $[0,1]^n$:

$$\mathbf{z} = \arg\min_{\mathbf{z}' \in [0,1]^n} S(\mathbf{z}', \mathbf{X}, \mathbf{Y}).$$

Решение (9) восстанавливается с помощью отсечения по порогу:

$$\mathbf{a} = [a_j]_{j=1}^n, \quad a_j = \begin{cases} 1, & z_j > \tau; \\ 0, & \text{в противном случае.} \end{cases}$$

Как только решение $\mathbf{a}$ задачи (9) получено, задача (3) принимает вид:

$$\mathcal{L}(\boldsymbol{\Theta}_{\mathcal{A}}, \mathbf{X}_{\mathcal{A}}, \mathbf{Y}) = \|\mathbf{Y} - \mathbf{X}_{\mathcal{A}}\boldsymbol{\Theta}_{\mathcal{A}}\|_2^2 \to \min_{\boldsymbol{\Theta}_{\mathcal{A}}}.$$

Если между столбцами матрицы исходных объектов $\mathbf{X}$ существует линейная зависимость, то решение задачи линейной регрессии

$$\|\boldsymbol{\upsilon} - \mathbf{X}\boldsymbol{\theta}\|_2^2 \to \min_{\boldsymbol{\theta} \in \mathbb{R}^n}.$$

оказывается неустойчивым. Методы выбора признаков находят подмножество $\mathcal{A} \in \{1, \ldots, n\}$ оптимальных столбцов матрицы $\mathbf{X}$.

Метод QPFS выбирает некоррелированные признаки, релевантные целевому вектору $\boldsymbol{\upsilon}$. Введем две функции: $\mathrm{Sim}(\mathbf{X})$ контролирует избыточность между признаками, $\mathrm{Rel}(\mathbf{X}, \boldsymbol{\upsilon})$ содержит релевантности между каждым признаком и целевым вектором.

Метод QPFS минимизирует следующую функцию ошибки

$$\underbrace{\mathbf{z}^{\intercal}\mathbf{Q}\mathbf{z}}_{\mathrm{Sim}} - \alpha \cdot \underbrace{\mathbf{b}^{\intercal}\mathbf{z}}_{\mathrm{Rel}} \to \min_{\substack{\mathbf{z} \in \mathbb{R}_+^n \\ \|\mathbf{z}\|_1 = 1}}. \tag{10}$$

Элементы матрицы парных взаимодействий $\mathbf{Q} \in \mathbb{R}^{n \times n}$ содержат коэффициенты попарного сходства между признаками. Вектор релевантностей признаков $\mathbf{b} \in \mathbb{R}^n$ выражает сходство между каждым признаком и целевым вектором $\boldsymbol{\upsilon}$. Нормированный вектор $\mathbf{z}$ отражает значимость каждого признака. Параметр α позволяет контролировать компромисс между Sim и Rel:

$$\alpha = \frac{\overline{\mathbf{Q}}}{\overline{\mathbf{Q}} + \overline{\mathbf{b}}}, \quad \text{где } \overline{\mathbf{Q}} = \mathrm{mean}(\mathbf{Q}), \ \ \overline{\mathbf{b}} = \mathrm{mean}(\mathbf{b}).$$

Для измерения сходства используется выборочный коэффициент корреляции Пирсона между парами признаков для функции Sim, и между признаками и целевым вектором для функции Rel:

$$\mathbf{Q} = \left[|\mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)|\right]_{i,j=1}^n, \quad \mathbf{b} = \left[|\mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\upsilon})|\right]_{i=1}^n.$$

Здесь

$$\operatorname{corr}(\boldsymbol{\chi}, \boldsymbol{\upsilon}) = \frac{\sum_{i=1}^{m}(\boldsymbol{\chi}_i - \overline{\boldsymbol{\chi}})(\boldsymbol{\upsilon}_i - \overline{\boldsymbol{\upsilon}})}{\sqrt{\sum_{i=1}^{m}(\boldsymbol{\chi}_i - \overline{\boldsymbol{\chi}})^2 \sum_{i=1}^{m}(\boldsymbol{\upsilon}_i - \overline{\boldsymbol{\upsilon}})^2}}.$$

В случае векторной целевой переменной компоненты целевой переменной могут коррелировать между собой. Предлагаются методы, учитывающие зависимости как в исходном, так и в целевом пространствах.

Агрегация релевантностей целевых векторов. Чтобы применить метод QPFS к векторному случаю ($r > 1$), релевантности признаков агрегируются по всем r компонентам целевой переменной. Вектор релевантностей $\mathbf{b}$ агрегируется по всем компонентам целевой переменной и определяется как

$$\mathbf{b} = \left[\sum_{k=1}^{r} |\operatorname{corr}(\boldsymbol{\chi}_i, \boldsymbol{\upsilon}_k)|\right]_{i=1}^{n}.$$

Недостатком такого подхода является отсутствие учёта зависимостей в столбцах матрицы $\mathbf{Y}$.

Симметричный учёт значимости признаков и целевых переменных. Добавим член $\operatorname{Sim}(\mathbf{Y})$ и изменим член $\operatorname{Rel}(\mathbf{X}, \mathbf{Y})$ следующим образом:

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\intercal}\mathbf{Q}_x\mathbf{z}_x}_{\operatorname{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\intercal}\mathbf{B}\mathbf{z}_y}_{\operatorname{Rel}(\mathbf{X},\mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\intercal}\mathbf{Q}_y\mathbf{z}_y}_{\operatorname{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n,\, \mathbf{1}_n^{\intercal}\mathbf{z}_x = 1 \\ \mathbf{z}_y \geqslant \mathbf{0}_r,\, \mathbf{1}_r^{\intercal}\mathbf{z}_y = 1}}. \tag{11}$$

Определим элементы матриц $\mathbf{Q}_x \in \mathbb{R}^{n\times n}$, $\mathbf{Q}_y \in \mathbb{R}^{r\times r}$ и $\mathbf{B} \in \mathbb{R}^{n\times r}$ следующим образом:

$$\mathbf{Q}_x = \left[|\operatorname{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)|\right]_{i,j=1}^{n}, \quad \mathbf{Q}_y = \left[|\operatorname{corr}(\boldsymbol{\upsilon}_i, \boldsymbol{\upsilon}_j)|\right]_{i,j=1}^{r}, \quad \mathbf{B} = \left[|\operatorname{corr}(\boldsymbol{\chi}_i, \boldsymbol{\upsilon}_j)|\right]_{\substack{i=1,\ldots,n \\ j=1,\ldots,r}}.$$

Вектор $\mathbf{z}_x$ содержит коэффициенты значимости признаков, $\mathbf{z}_y$ — коэффициенты значимости целевых векторов.

Утверждение 9. Баланс между $\operatorname{Sim}(\mathbf{X})$, $\operatorname{Rel}(\mathbf{X}, \mathbf{Y})$ и $\operatorname{Sim}(\mathbf{Y})$ в задаче (11) достигается при:

$$\alpha_1 \propto \overline{\mathbf{Q}}_y\overline{\mathbf{B}}, \quad \alpha_2 \propto \overline{\mathbf{Q}}_x\overline{\mathbf{Q}}_y, \quad \alpha_3 \propto \overline{\mathbf{Q}}_x\overline{\mathbf{B}}.$$

Здесь $\overline{\mathbf{Q}}_x$, $\overline{\mathbf{B}}$ и $\overline{\mathbf{Q}}_y$ — средние значения соответствующих матриц $\mathbf{Q}_x$, $\mathbf{B}$ и $\mathbf{Q}_y$ членов $\operatorname{Sim}(\mathbf{X})$, $\operatorname{Rel}(\mathbf{X}, \mathbf{Y})$ и $\operatorname{Sim}(\mathbf{Y})$.

Минимаксная постановка задачи выбора признаков. Сформулируем две взаимосвязанные задачи:

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\intercal}\mathbf{Q}_x\mathbf{z}_x}_{\operatorname{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\intercal}\mathbf{B}\mathbf{z}_y}_{\operatorname{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n, \\ \mathbf{1}_n^{\intercal}\mathbf{z}_x = 1}}, \tag{12}$$

$$\alpha_3 \cdot \underbrace{\mathbf{z}_y^{\intercal}\mathbf{Q}_y\mathbf{z}_y}_{\operatorname{Sim}(\mathbf{Y})} + \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\intercal}\mathbf{B}\mathbf{z}_y}_{\operatorname{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r, \\ \mathbf{1}_r^{\intercal}\mathbf{z}_y = 1}}. \tag{13}$$

Задачи (12) и (13) объединяются в совместную минимакс или максмин постановку

$$\min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^\mathsf{T}\mathbf{z}_x=1}} \max_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^\mathsf{T}\mathbf{z}_y=1}} f(\mathbf{z}_x, \mathbf{z}_y), \quad \left(\text{или } \max_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^\mathsf{T}\mathbf{z}_y=1}} \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^\mathsf{T}\mathbf{z}_x=1}} f(\mathbf{z}_x, \mathbf{z}_y) \right), \tag{14}$$

где

$$f(\mathbf{z}_x, \mathbf{z}_y) = \alpha_1 \cdot \underbrace{\mathbf{z}_x^\mathsf{T}\mathbf{Q}_x\mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^\mathsf{T}\mathbf{B}\mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} - \alpha_3 \cdot \underbrace{\mathbf{z}_y^\mathsf{T}\mathbf{Q}_y\mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})}.$$

Теорема 2. Для положительно определенной матрицы $\mathbf{Q}_x$ и $\mathbf{Q}_y$, максмин и минимакс задачи (14) имеют одинаковое оптимальное значение.

Утверждение 10. Минимаксная задача (14) эквивалентна задаче квадратичного программирования с $n + r + 1$ переменными.

Несимметричный учёт значимостей признаков и целевых переменных. Добавим линейный член $\mathbf{b}^\mathsf{T}\mathbf{z}_y$ в член $\mathrm{Rel}(\mathbf{X}, \mathbf{Y})$ следующим образом:

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^\mathsf{T}\mathbf{Q}_x\mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\left(\mathbf{z}_x^\mathsf{T}\mathbf{B}\mathbf{z}_y - \mathbf{b}^\mathsf{T}\mathbf{z}_y\right)}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{z}_y^\mathsf{T}\mathbf{Q}_y\mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n,\, \mathbf{1}_n^\mathsf{T}\mathbf{z}_x=1 \\ \mathbf{z}_y \geqslant \mathbf{0}_r,\, \mathbf{1}_r^\mathsf{T}\mathbf{z}_y=1}}. \tag{15}$$

Утверждение 11. Пусть вектор $\mathbf{b}$ равен

$$b_j = \max_{i=1,\dots n} [\mathbf{B}]_{i,j}.$$

Тогда значение коэффициентов значимостей вектора $\mathbf{z}_y$ будут неотрицательными в $\mathrm{Rel}(\mathbf{X}, \mathbf{Y})$ для задачи (15).

Утверждение 12. Баланс между членами $\mathrm{Sim}(\mathbf{X})$, $\mathrm{Rel}(\mathbf{X}, \mathbf{Y})$ и $\mathrm{Rel}(\mathbf{X}, \mathbf{Y})$ для задачи (15) достигается при следующих коэффициентах:

$$\alpha_1 \propto \overline{\mathbf{Q}}_y \left(\overline{\mathbf{b}} - \overline{\mathbf{B}}\right), \quad \alpha_2 \propto \overline{\mathbf{Q}}_x \overline{\mathbf{Q}}_y, \quad \alpha_3 \propto \overline{\mathbf{Q}}_x \overline{\mathbf{B}}.$$

Теорема 3. В случае скалярной целевой переменной ($r = 1$) предлагаемые методы выбора признаков SymImp (11), MinMax (14), AsymImp (15) совпадают с оригинальным методом QPFS (10).

Таблица 1 демонстрирует основные идеи и функции ошибок для каждого метода. RelAgg является базовой стратегией и не учитывает корреляции в целевом пространстве. SymImp штрафует попарные корреляции между целевыми векторами. MinMax более чувствителен к целевым векторам, которые трудно предсказать. Стратегия Asymimp добавляет линейный член к функции SymImp, чтобы сделать вклад признаков и целевых векторов асимметричным.

В **главе 4** предлагается метод выбора активных параметров модели с использованием метода выбора признаков с помощью квадратичного программирования. Приводится анализ параметров модели, которые не находятся в оптимуме.

Таблица 1: Обзор предлагаемых обобщений метода QPFS для векторной целевой переменной

Метод	Идея	Функция ошибки $S(\mathbf{z}\|\mathbf{X},\mathbf{Y})$
RelAgg	$\min\big[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\big]$	$\min\limits_{\mathbf{z}_x}\big[(1-\alpha)\cdot\mathbf{z}_x^\intercal\mathbf{Q}_x\mathbf{z}_x - \alpha\cdot\mathbf{z}_x^\intercal\mathbf{B}\mathbf{1}_r\big]$
SymImp	$\min\big[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\big]$	$\min\limits_{\mathbf{z}_x,\,\mathbf{z}_y}\Big[\alpha_1\cdot\mathbf{z}_x^\intercal\mathbf{Q}_x\mathbf{z}_x - \alpha_2\cdot\mathbf{z}_x^\intercal\mathbf{B}\mathbf{z}_y + \alpha_3\cdot\mathbf{z}_y^\intercal\mathbf{Q}_y\mathbf{z}_y\Big]$
MinMax	$\min\big[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\big]$ $\max\big[\mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\big]$	$\min\limits_{\mathbf{z}_x}\max\limits_{\mathbf{z}_y}\big[\alpha_1\cdot\mathbf{z}_x^\intercal\mathbf{Q}_x\mathbf{z}_x - \alpha_2\cdot\mathbf{z}_x^\intercal\mathbf{B}\mathbf{z}_y - \alpha_3\cdot\mathbf{z}_y^\intercal\mathbf{Q}_y\mathbf{z}_y\big]$
AsymImp	$\min\big[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\big]$ $\max\big[\mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\big]$	$\min\limits_{\mathbf{z}_x,\mathbf{z}_y}\big[\alpha_1\cdot\mathbf{z}_x^\intercal\mathbf{Q}_x\mathbf{z}_x - \alpha_2\cdot\Big(\mathbf{z}_x^\intercal\mathbf{B}\mathbf{z}_y - \mathbf{b}^\intercal\mathbf{z}_y\Big) + \alpha_3\cdot\mathbf{z}_y^\intercal\mathbf{Q}_y\mathbf{z}_y\big]$

Модель $f(\mathbf{x},\boldsymbol{\theta})$ с параметрами $\boldsymbol{\theta}\in\mathbb{R}^p$ предсказывает целевой объект $y\in\mathbb{Y}$ по исходному объекту $\mathbf{x}\in\mathbb{R}^n$. Пространство $\mathbb{Y}$ представляет собой бинарные метки классов $\{0,1\}$ для задачи двухклассовой классификации и $\mathbb{R}$ для задачи регрессии. Параметры $\boldsymbol{\theta}$ вычисляются минимизацией функции ошибки:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p}\mathcal{L}(\boldsymbol{\theta},\mathbf{X},\mathbf{y}). \tag{16}$$

В качестве функции ошибки $\mathcal{L}(\boldsymbol{\theta},\mathbf{X},\mathbf{y})$ рассматриваются квадратичная ошибка для задачи регрессии:

$$\mathcal{L}(\boldsymbol{\theta},\mathbf{X},\mathbf{y}) = \frac{1}{2}\|\mathbf{y}-\mathbf{f}(\mathbf{X},\boldsymbol{\theta})\|_2^2 = \frac{1}{2}\sum_{i=1}^{m}\big(y_i - f(\mathbf{x}_i,\boldsymbol{\theta})\big)^2, \tag{17}$$

и функция кросс-энтропии для задачи бинарной классификации:

$$\mathcal{L}(\boldsymbol{\theta},\mathbf{X},\mathbf{y}) = \sum_{i=1}^{m}\big[y_i\log f(\mathbf{x}_i,\boldsymbol{\theta}) + (1-y_i)\log\big(1-f(\mathbf{x}_i,\boldsymbol{\theta})\big)\big]. \tag{18}$$

Для выбора вектора обновлений $\Delta\boldsymbol{\theta}$ используется метод оптимизации Ньютона .

Метод Ньютона нестабилен и вычислительно сложен. В работе предлагается стабильный метод Ньютона. Перед шагом градиента предлагается выбрать подмножество активных параметров модели, которые оказывают наибольшее влияние на функцию ошибки $\mathcal{L}(\boldsymbol{\theta},\mathbf{X},\mathbf{y})$.

Определение 11. Параметр θ_j для модели $f(\mathbf{x},\boldsymbol{\theta})$ является *активным*, если $\mathbf{J}^\intercal(\mathbf{f}(\mathbf{x},\boldsymbol{\theta})-\mathbf{y})\neq 0$.

Обновление параметров производится только для отобранного множества индексов $\mathcal{A} = \big\{j : a_j = 1, \mathbf{a}\in\{0,1\}^p\big\}$

$$\boldsymbol{\theta}_{\mathcal{A}}^k = \boldsymbol{\theta}_{\mathcal{A}}^{k-1} + \Delta\boldsymbol{\theta}_{\mathcal{A}}^{k-1}, \quad \boldsymbol{\theta}_{\mathcal{A}} = \{\theta_j : j\in\mathcal{A}\},$$
$$\boldsymbol{\theta}_{\bar{\mathcal{A}}}^k = \boldsymbol{\theta}_{\bar{\mathcal{A}}}^{k-1}, \quad \boldsymbol{\theta}_{\bar{\mathcal{A}}} = \{\theta_j : j\notin\mathcal{A}\}.$$

Чтобы выбрать оптимальное подмножество индексов $\mathcal{A}$, из всех возможных $2^p - 1$ подмножеств, используется метод QPFS. Метод выбирает подмножество параметров $\mathbf{a}$ для вектора обновлений $\Delta\boldsymbol{\theta}$, которые оказывают наибольшее влияние на вектор остатков и являются попарно независимыми:

$$\mathbf{a} = \arg\max_{\mathbf{a}' \in \{1,0\}^p} S(\mathbf{a}', \mathbf{X}, \mathbf{y}, \boldsymbol{\theta}) \Leftrightarrow \arg\min_{\mathbf{a} \in \mathbb{R}^p_+,\, \|\mathbf{a}\|_1 = 1} \left[\mathbf{a}^\intercal \mathbf{Q} \mathbf{a} - \alpha \cdot \mathbf{b}^\intercal \mathbf{a}\right].$$

Метод Ньютона использует условие оптимизации первого порядка для задачи (16) и линеаризует градиент $S(\boldsymbol{\theta})$

$$\nabla S(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}) = \nabla S(\boldsymbol{\theta}) + \mathbf{H} \cdot \Delta\boldsymbol{\theta} = 0, \quad \Delta\boldsymbol{\theta} = -\mathbf{H}^{-1} \nabla S(\boldsymbol{\theta}).$$

где $\mathbf{H} = \nabla^2 S(\boldsymbol{\theta})$ является матрицей Гессиана функции ошибки $S(\boldsymbol{\theta})$.

Теорема 4. Пусть модель $f(\mathbf{x}, \boldsymbol{\theta})$ близка к линейной в окрестности точки $\boldsymbol{\theta} + \Delta\boldsymbol{\theta}$

$$\mathbf{f}(\mathbf{X}, \boldsymbol{\theta} + \Delta\boldsymbol{\theta}) \approx \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}) + \mathbf{J} \cdot \Delta\boldsymbol{\theta},$$

где $\mathbf{J} \in \mathbb{R}^{m \times p}$ является матрицей Якоби. Тогда вектор обновления $\Delta\boldsymbol{\theta}$ для функции ошибки (17) является решением задачи линейной регрессии

$$\|\mathbf{e} - \mathbf{F}\Delta\boldsymbol{\theta}\|_2^2 \to \min_{\Delta\boldsymbol{\theta} \in \mathbb{R}^p}, \tag{19}$$

где $\mathbf{e} = \mathbf{f} - \mathbf{y}$ и $\mathbf{F} = \mathbf{J}$.

Теорема 5. Рассмотрим модель логистической регрессии вида $f(\mathbf{x}, \boldsymbol{\theta}) = \sigma(\mathbf{x}^\intercal \boldsymbol{\theta})$ с сигмоидной функцией активации $\sigma(\cdot)$. Вектор обновлений $\Delta\boldsymbol{\theta}$ для функции ошибки (18) является решением задачи линейной регрессии

$$\|\mathbf{e} - \mathbf{F}\Delta\boldsymbol{\theta}\|_2^2 \to \min_{\Delta\boldsymbol{\theta} \in \mathbb{R}^p}, \tag{20}$$

где $\mathbf{e} = \mathbf{R}^{-1/2}(\mathbf{y} - \mathbf{f})$ и $\mathbf{F} = \mathbf{R}^{1/2}\mathbf{X}$.

Предлагается адаптация метода QPFS для решения задач (19) и (20). Матрица парных взаимодействий $\mathbf{Q}$ и вектор релевантностей $\mathbf{b}$ имеют вид

$$\mathbf{Q} = \mathrm{Sim}(\mathbf{F}), \quad \mathbf{b} = \mathrm{Rel}(\mathbf{F}, \mathbf{e}).$$

Утверждение 13. В оптимальной точке $\boldsymbol{\theta}^*$ вектор релевантностей $\mathbf{b} = \mathrm{Rel}(\mathbf{F}, \mathbf{e})$ равен нулю.

В **главе 5** ставится задача метрического обучения как поиск оптимальной метрики в целевом пространстве. При использовании в качестве функции ошибки квадратичной ошибки предполагается, что целевое пространство является евклидовым. Данное предположение не всегда является адекватным. Рассматриваются задачи кластеризации и классификации множества временных рядов.

Пусть $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]^\intercal \in \mathbb{R}^{m \times n}$ — матрица исходных объектов. Требуется выявить кластерную структуру данных и разбить множество объектов $\mathbf{X}$ на множество непересекающихся кластеров $\mathbb{Y} = \{1, \ldots, K\}$, т. е. построить отображение $f : \mathbb{R}^n \to \mathbb{Y}$. Обозначим $y_i = f(\mathbf{x}_i)$, $y_i \in \mathbb{Y}$ — метка кластера объекта $\mathbf{x}_i$.

Определение 12. *Центроидом* класса $e \in \mathbb{Y}$ множества объектов $\mathbf{X}_e$ по расстоянию ρ назовем вектор $\mathbf{c}_e \in \mathbb{R}^n$ такой, что

$$\mathbf{c}_e = \underset{\mathbf{c}\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{\mathbf{x}_i\in\mathbf{X}_e} \rho(\mathbf{x}_i,\mathbf{c}). \tag{21}$$

Здесь $\mathbf{X}_e = \{\mathbf{x}_i : i = 1,\ldots,m,\ y_i = e\}$ — множество объектов выборки, принадлежащих одному классу $e \in \mathbb{Y}$.

В случае использования в качестве расстояния ρ евклидовой метрики формула (21) принимает вид

$$\mathbf{c} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i, \quad \mathbf{c}_e = \frac{\sum_{i=1}^{m}[y_i = y_e]\mathbf{x}_i}{\sum_{i=1}^{m}[y_i = y_e]}.$$

Здесь вектор $\mathbf{c}$ является центроидом всего множества объектов $\mathbf{X}$.

Введем на множестве объектов $\mathbf{X}$ расстояние Махаланобиса

$$d_{\mathbf{A}}(\mathbf{x}_i,\mathbf{x}_j) = \sqrt{(\mathbf{x}_i-\mathbf{x}_j)^{\top}\mathbf{A}^{-1}(\mathbf{x}_i-\mathbf{x}_j)}, \tag{22}$$

где *матрица трансформаций* $\mathbf{A} \in \mathbb{R}^{n\times n}$ является симметричной и неотрицательно определенной ($\mathbf{A}^{\intercal} = \mathbf{A}$, $\mathbf{A} \succeq 0$). Зададим в качестве матрицы трансформации матрицу выборочной ковариации

$$\mathbf{A} = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i-\mathbf{c})(\mathbf{x}_i-\mathbf{c})^{\intercal}. \tag{23}$$

Функцией ошибки кластеризации назовем межкластерное расстояние:

$$\mathcal{L}\big(\{\mathbf{c}_j\}_{j=1}^{K},\mathbf{X},\mathbf{y}\big) = -\sum_{j=1}^{K} N_j d^2_{\mathbf{A}}(\mathbf{c}_j,\mathbf{c}), \tag{24}$$

где $N_j = \sum_{i=1}^{m}[y_i = y_j]$ — число объектов в кластере j.

Поставим задачу кластеризации как задачу минимизации функции ошибки (24)

$$\mathcal{L}\big(\{\mathbf{c}_j\}_{j=1}^{K},\mathbf{X},\mathbf{y}\big) \to \min_{\mathbf{c}_j\in\mathbb{R}^{\intercal}}. \tag{25}$$

Найдем такую матрицу $\mathbf{A}$, для которой функционал качества принимает максимальное значение:

$$\mathbf{A}^* = \underset{\mathbf{A}\in\mathbb{R}^{n\times n}}{\arg\min}\, S\big(\{\mathbf{c}_j^*\}_{j=1}^{K},\mathbf{X},\mathbf{y}\big), \tag{26}$$

где $\{\mathbf{c}_j^*\}_{j=1}^{K}$ — решение задачи кластеризации (25).

Для решения задач (25), (26) используется алгоритм адаптивного метрического обучения. Предлагается понизить размерность пространства объектов $\mathbf{X}$

с помощью линейного ортогонального преобразования $\mathbf{P} \in \mathbb{R}^{l \times n}$, $\mathbf{P}^\intercal \mathbf{P} = \mathbf{I}$, где новая размерность $l < n$

$$\mathbf{t}_i = \mathbf{P}\mathbf{x}_i \in \mathbb{R}^l, \quad i = 1, \ldots, m.$$

Расстояния между объектами вычисляются по формуле (22), где в качестве матрицы $\hat{\mathbf{A}}$ используется матрица ковариаций (23) множества объектов $\{\hat{\mathbf{x}}_i\}_{i=1}^m$

$$\hat{\mathbf{A}} = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{t}_i - \hat{\mathbf{c}})(\mathbf{t}_i - \hat{\mathbf{c}})^\intercal = \frac{1}{m}\sum_{i=1}^{m}\mathbf{P}(\mathbf{x}_i - \mathbf{c})(\mathbf{x}_i - \mathbf{c})^\intercal\mathbf{P}^\intercal = \mathbf{P}\mathbf{A}\mathbf{P}^\intercal.$$

Определение 13. *Индикаторной матрицей* назовем матрицу $\mathbf{Y} \in \mathbb{R}^{m \times K}$, где $y_{ij} = [f(\mathbf{x}_i) = y_j]$.

Определение 14. *Взвешенной индикаторной матрицей* назовем матрицу $\mathbf{L} = \mathbf{Y}(\mathbf{Y}^\intercal\mathbf{Y})^{-1/2} \in \mathbb{R}^{m \times K}$, элементы которой равны:

$$l_{ij} = \begin{cases} \dfrac{1}{\sqrt{N_j}}, & \text{если } f(\mathbf{x}_i) = y_j; \\ 0, & \text{если } f(\mathbf{x}_i) \neq y_j. \end{cases}$$

Задача кластеризации (25) и задача метрического обучения (26) сводятся к общей задаче минимизации функции ошибки

$$\mathcal{L} = -\frac{1}{m}\mathrm{trace}(\mathbf{L}^\intercal\mathbf{X}^\intercal\mathbf{P}^\intercal\hat{\mathbf{A}}^{-1}\mathbf{P}\mathbf{X}\mathbf{L}) = \\ = -\frac{1}{m}\mathrm{trace}(\mathbf{L}^\intercal\mathbf{X}^\intercal\mathbf{P}^\intercal(\mathbf{P}\mathbf{A}\mathbf{P}^\intercal)^{-1}\mathbf{P}\mathbf{X}\mathbf{L}) \to \min_{\mathbf{P},\mathbf{L}}. \quad (27)$$

Для решения задачи (27) используется EM алгоритм.

Оптимизация матрицы P с фиксированной матрицей L. Переформулируем задачу (27) следующим образом:

$$\mathcal{L} = -\frac{1}{m}\mathrm{trace}(\mathbf{L}^\intercal\mathbf{X}^\intercal\mathbf{P}^\intercal(\mathbf{P}\mathbf{A}\mathbf{P}^\intercal)^{-1}\mathbf{P}\mathbf{X}\mathbf{L}) = -\frac{1}{m}\mathrm{trace}\big((\mathbf{P}\mathbf{A}\mathbf{P}^\intercal)^{-1}\mathbf{P}\mathbf{X}\mathbf{L}\mathbf{L}^T\mathbf{X}^\intercal\mathbf{P}^\intercal\big).$$

Утверждение 14. Обозначим $\mathbf{B} = \mathbf{X}\mathbf{L}\mathbf{L}^\intercal\mathbf{X}^\intercal$. Обозначим через $\mathbf{P} = [\mathbf{v}_1, \ldots, \mathbf{v}_r]^\mathbf{T}$ матрицу, состоящую из r собственных векторов матрицы $\mathbf{A}^{-1}\mathbf{B}$, отвечающих наибольшим собственным значениям. Тогда решением (27) является ортогональная матрица, полученная QR-разложением матрицы $\mathbf{P}^\intercal$.

Оптимизация матрицы L с фиксированной матрицей P. Обозначим $\hat{\mathbf{K}} = (1/N)\mathbf{X}^\intercal\mathbf{P}^\intercal\hat{\mathbf{A}}^{-1}\mathbf{P}\mathbf{X}$.

При фиксированной матрице $\mathbf{P}$ задача (27) принимает вид:

$$\mathrm{trace}(\mathbf{L}^\top\hat{\mathbf{K}}\mathbf{L}) \to \max_{\mathbf{L}\in\mathbf{R}^{m\times r}}.$$

Для нахождения оптимального соответствия между временными рядами при решении задачи классификации предлагается процедура выравнивания временных рядов. Пусть исходный объект $\mathbf{x}_i \in \mathbb{R}^n$ — временной ряд, последовательность измерений некоторой исследуемой величины в различные моменты времени. Пусть задана выборка $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$ — множество объектов с известными метками классов $y_i \in \mathbb{Y}$, где $\mathbb{Y} = \{1, \ldots, K\}$ — множество меток классов.

Требуется построить точную, простую, устойчивую модель классификации $a : \mathbb{R}^n \to \mathbb{Y}$. Данную модель представим в виде суперпозиции

$$a(\mathbf{x}) = b \circ \mathbf{f} \circ G\big(\mathbf{x}, \{\mathbf{c}_e\}_{e=1}^K\big),$$

где G — процедура выравнивания временных рядов относительно центроидов классов $\{\mathbf{c}_e\}_{e=1}^K$, $\mathbf{f}$ — алгоритм метрического обучения, b — алгоритм многоклассовой классификации.

Для нахождения центроида предлагается в качестве расстояния ρ между временными рядами использовать путь наименьшей стоимости [**?**, **?**], найденный методом динамической трансформации времени.

Метрическое обучение. Введем на множестве выравненных временных рядов расстояние Махаланобиса $d_{\mathbf{A}}$ 22. Представим матрицу трансформации $\mathbf{A}$ в виде разложения $\mathbf{A}^{-1} = \mathbf{L}^\intercal\mathbf{L}$. Матрица $\mathbf{L} \in \mathbb{R}^{p\times n}$ — матрица линейного преобразования, где p задает размерность преобразованного пространства.

Расстояние $d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j)$ есть евклидово расстояние между $\mathbf{L}\mathbf{x}_i$ и $\mathbf{L}\mathbf{x}_j$:

$$d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^\intercal\mathbf{L}^\intercal\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)} = \\ = \sqrt{(\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j))^\intercal(\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j))} = \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|_2.$$

В качестве алгоритма метрического обучения в данной работе был выбран алгоритм LMNN. Данный алгоритм сочетает в себе идеи метода k ближайших соседей. Первая идея заключается в минимизации расстояний между k ближайшими объектами, находящимися в одном классе:

$$Q_1(\mathbf{L}) = \sum_{j \rightsquigarrow i} \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 \to \min_{\mathbf{L}},$$

где $j \rightsquigarrow i$ означает, что $\mathbf{x}_j$ является одним из k ближайших соседей для $\mathbf{x}_i$. Вторая идея состоит в максимизации расстояния между каждым объектом и его объектами-нарушителями.

Определение 15. *Объектом-нарушителем* для $\mathbf{x}_i$ назовем объект $\mathbf{x}_l$ такой, что

$$\|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)\|^2 \leqslant \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 + 1, \quad \text{где } j \rightsquigarrow i.$$

Таким образом, необходимо минимизировать следующий функционал:

$$Q_2(\mathbf{L}) = \sum_{j \rightsquigarrow i} \sum_{l} [y_i \neq y_l] \left[1 + \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_j)\|^2 - \|\mathbf{L}(\mathbf{x}_i - \mathbf{x}_l)\|^2\right]_+ \to \min_{\mathbf{L}}.$$

Задача метрического обучения состоит в нахождении линейного преобразования $\mathbf{f}(\mathbf{x}) = \mathbf{L}\mathbf{x}$, то есть нахождении матрицы $\mathbf{L}$ в виде решения оптимизационной задачи

$$Q(\mathbf{L}) = \mu Q_1(\mathbf{L}) + (1 - \mu) Q_2(\mathbf{L}) \to \min_{\mathbf{L}}.$$

В **главе 6** ставится задача порождения информативного признакового пространства. Временные ряды акселерометра образуют множество $\mathcal{S}$ сегментов $\mathbf{s} = [x_1, \ldots, x_T]^\intercal$ фиксированной длины T. Необходимо построить модель классификации $a : \mathbb{R}^T \to \mathbb{Y}$, которая будет ставить в соответствие каждому сегменту из множества $\mathcal{S}$ метку класса из конечного множества $\mathbb{Y} = \{1, \ldots, K\}$. Обозначим за $\mathcal{D} = \{(\mathbf{s}_i, y_i)\}_{i=1}^m$ исходную выборку, где $\mathbf{s}_i \in \mathcal{S}$ и $y_i = a(\mathbf{s}_i) \in \mathbb{Y}$.

В работе предлагается построить модель a в виде суперпозиции $a = f \circ \mathbf{g}$.

Определение 16. *Порождающей функцией* будем называть функцию $\mathbf{g} : \mathbb{R}^T \to \mathbb{X}$, отображающую исходные временные ряды $\mathbf{s}$ из пространства $\mathbb{R}^T$ в признаковое пространство $\mathbb{X} \subset \mathbb{R}^n$.

Имея порождающую функцию $\mathbf{g}$, преобразуем исходную выборку в $\mathcal{D}_\mathbb{X} = \{(\mathbf{x}_i, y_i)\}_{i=1}^m$, где $\mathbf{x}_i = \mathbf{g}(\mathbf{s}_i) \in \mathbb{X}$. Здесь $\mathbf{x}_i$ является исходным объектом.

Модель классификации $f = f(\mathbf{x}, \boldsymbol{\theta})$ является параметрической функцией с вектором параметров $\boldsymbol{\theta}$. Оптимальные параметры $\boldsymbol{\theta}^*$ определяются оптимизацией функции ошибки классификации

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mathcal{D}_\mathbb{X}, \boldsymbol{\mu}).$$

Экспертные функции. Экспертные функции — это некоторые статистики g_j, где $g_j : \mathbb{R}^T \to \mathbb{R}$. Признаковым описанием $\mathbf{g}(\mathbf{s})$ объекта s являются значения заданных экспертных статистик для данного объекта

$$\mathbf{x} = \mathbf{g}(\mathbf{s}) = [g_1(\mathbf{s}), \ldots, g_n(\mathbf{s})]^\intercal.$$

Авторегрессионная модель. Авторегрессионная модель порядка n использует параметрическую модель для аппроксимации временного ряда $\mathbf{s}$:

$$x_t = w_0 + \sum_{j=1}^{n-1} w_j x_{t-j} + \varepsilon_t.$$

Оптимальные параметры $\mathbf{w}^*$ авторегрессионной модели используются как признаки $\mathbf{g}(\mathbf{s})$:

$$\mathbf{x} = \mathbf{g}(\mathbf{s}) = \mathbf{w}^* = \arg\min_{\mathbf{w} \in \mathbb{R}^n} \left(\sum_{t=n}^{T} \|x_t - \hat{x}_t\|^2 \right).$$

Таблица 2: Примеры экспертных порождающих функций

Описание	**Формула**
Mean	$\bar{s} = \frac{1}{T}\sum_{t=1}^{T} s_t$
Standard deviation	$\sqrt{\frac{1}{T}\sum_{t=1}^{T}(s_t - \bar{s})^2}$
Mean absolute deviation	$\frac{1}{T}\sum_{t=1}^{T}\lvert s_t - \bar{s}\rvert$
Distribution	Histogram values with 10 bins

Анализ сингулярного спектра. Для каждого временного ряда $\mathbf{s}$ из исходной выборки $\mathcal{D}$ строится траекторная матрица:

$$\mathbf{S} = \begin{pmatrix} s_1 & s_2 & \dots & s_n \\ s_2 & s_3 & \dots & s_{n+1} \\ \dots & \dots & \dots & \dots \\ s_{T-n+1} & s_{T-n+2} & \dots & s_T \end{pmatrix}.$$

Сингулярное разложение матрицы $\mathbf{S}^\intercal\mathbf{S}$:

$$\mathbf{S}^\intercal\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\intercal.$$

Признаковое описание объекта $\mathbf{s}$ задаётся спектром матрицы $\mathbf{S}^\intercal\mathbf{S}$:

$$\mathbf{x} = \mathbf{g}(\mathbf{s}) = [\lambda_1, \dots, \lambda_n]^\intercal.$$

Аппроксимация сплайнами. Предполагается, что узлы сплайна $\{\xi_\ell\}_{\ell=0}^{M}$ равномерно распределены по временной оси. Кусочные модели, построенные на отрезках $[\xi_{\ell-1}; \xi_\ell]$, заданы коэффициентами $\{\mathbf{w}_\ell\}_{\ell=1}^{M}$. Обозначим каждый отрезок-сегмент $p_i(t)$ $i = 1, \dots, M$ и весь сплайн $S(t)$:

$$S(t) = \begin{cases} p_1(t) = w_{10} + w_{11}t + w_{12}t^2 + w_{13}t^3, & t \in [\xi_0, \xi_1], \\ p_2(t) = w_{20} + w_{21}t + w_{22}t^2 + w_{23}t^3, & t \in [\xi_1, \xi_2], \\ \dots & \dots \\ p_M(t) = w_{L0} + w_{M1}t + w_{M2}t^2 + w_{M3}t^3, & t \in [\xi_{M-1}, \xi_M], \end{cases}$$

$$\begin{aligned} S(\xi_t) = x_t, \quad t = 0, \dots, M, \\ p_i'(\xi_i) = p_{i+1}'(\xi_i), p_i''(\xi_i) = p_{i+1}''(\xi_i), \quad i = 1, \dots, M-1, \\ p_i(\xi_{i-1}) = x_{i-1}, p_i(\xi_i) = x_i, \quad i = 1, \dots, M. \end{aligned}$$

Объединение всех параметров сплайна задаёт признаковое описание временного ряда:

$$\mathbf{x} = \mathbf{g}(\mathbf{s}) = [\mathbf{w}_1, \dots, \mathbf{w}_M]^\intercal.$$

В заключении представлены основные результаты диссертационной работы.

1. Исследована проблема снижения размерности сигналов в коррелированных пространствах высокой размерности. Предложены методы декодирования сигналов, учитывающие зависимости как в исходном, так и в целевом пространстве сигналов.
2. Доказаны теоремы об оптимальности предлагаемых методов декодирования сигналов. Предлагаемые методы выбирают согласованные модели в случае избыточной размерности описания данных.
3. Предложены методы выбора признаков, учитывающие зависимости как в исходном, так и в целевом пространстве. Предложенные методы доставляют устойчивые и адекватные решения в пространствах высокой размерности.
4. Предложены нелинейные методы согласования скрытых пространств для данных со сложноорганизованной целевой переменной.
5. Предложен ряд моделей для прогнозирования гетерогенных наборов сигналов для задачи построения нейрокомпьютерных интерфейсов.

Публикации соискателя по теме диссертации

Публикации в журналах из списка ВАК.

1. Исаченко Р. В., Стрижов В. В. Метрическое обучение в задачах мультиклассовой классификации временных рядов // Информатика и её применения, 2016. Т. 10. № 2. С. 48–57.
2. Isachenko R. et al. Feature Generation for Physical Activity Classification // Artificial Intelligence and Decision Making, 2018. № 3. С. 20–27.
3. Isachenko R. V., Strijov V. V. Quadratic programming optimization with feature selection for nonlinear models // Lobachevskii Journal of Mathematics, 2018. Т. 39. № 9. С. 1179–1187.
4. Isachenko R. V., Vladimirova M. R., Strijov V. V. Dimensionality Reduction for Time Series Decoding and Forecasting Problems //DEStech Transactions on Computer Science and Engineering, 2018. №. optim.
5. Исаченко Р.В., Яушев Ф.Ю., Стрижов В.В. Модели согласования скрытого пространства в задаче прогнозирования // Системы и средства информатики, 2021. Т. 31. № 1.

Прочие публикации.

6. Исаченко Р. В., Катруца А. М. Метрическое обучение и снижение размерности пространства в задачах кластеризации // Машинное обучение и анализ данных, 2016. Т. 2. № 1. С. 17–25.